RNA-Seq

Q1. What is meant by RNA-Seq ?

A1. From Illumina [1] : *"RNA-Seq is a powerful sequencing-based method that enables researchers to discover, profile, and quantify RNA transcripts across the entire transcriptome. Because the method does not require probes or primers, the generated data are completely unbiased, allowing for hypothesis-free experimental design. ... Beyond gene expression analysis, RNA-Seq can identify novel transcripts, novel isoforms, alternative splice sites, allele-specific expression, and rare transcripts in a single experiment."*

Q2. What material should I send to be analyzed by RNA-Seq, total RNA or polyA(+) RNA ?

A2. Generally, we start with 1 μg of total RNA, then perform polyA(+) selection before making a library for sequencing. If instead polyA(+)-selected RNA is provided to NISC, then the investigator needs to provide at least 20 ng. As an alternative procedure, in this case to retain non-polyadenylated RNAs, NISC can deplete rRNA transcripts through selective hybridization, and then sequence the pool of residual RNAs. One ug of total RNA is required.

Q3. What data are returned by NISC ?

A3. Typically, NISC returns to the investigator a file containing basecalls and quality scores. The investigator is expected to provide data analyses; this is not offered by NISC. Commercial software is available from both Illumina and third party vendors; e.g., see

http://illumina.com/pagesnrn.ilmn?ID=229.

The data files can become quite large. Each HiSeq2000 lane can yield up to 280 million paired-end reads. For efficiency, a lane typically will contain a pool of barcoded samples, so demultiplexing is part of data processing. Quality of input nucleic acid will greatly influence the actual amount of quality sequences recovered. Poorly annotated genomes can make data analysis significantly more difficult.

Q4. How many times will a given sequence be found in the total data set (coverage) ?

A4. In the simplest case, the depth of coverage is a ratio of bases generated (number of fragments read × length of read) divided by bases of sequence targeted. If the entire transcriptome (human coding potential) were targeted (30 Mb), then one-third lane of 100b single-end reads are needed for 150-X coverage. This is a good starting point, but as with any experimental procedure, the read depth for any given sequence will vary from

the average. One can determine empirically through a pilot experiment whether more or fewer reads are needed to sufficiently cover a region of interest.

Q5. Is RNA-Seq equivalent to a microarray analysis of the transcriptome ?

A5. A microarray experiment retrieves RNAs of known sequence from a sample through hybridization to defined oligonucleotides bound to a solid surface. Fluorescence signal intensity from the hybrids indicates the relative amount of each of these RNAs in the original sample. This method has a 3-log dynamic range and is faster and cheaper than other methods of measuring gene expression.

An RNA-Seq experiment generates sequence reads for every RNA present in a sample. This method has at least 5-log dynamic range, and shows better quantitation and reproducibility than microarrays. Low abundance molecules are detected; if needed, detection limit is lowered by simply generating more sequencing reads. Unlike microarrays where analysis is limited to the set of known oligonucleotide sequences bound to the solid support, RNA-Seq permits discovery of novel transcripts and splice variants, distinguishes between closely-related sequences and provides for quantitation of expression between alleles. [2]

Q6. How many lanes of a flowcell are used for a mammalian cell RNA-Seq analysis?

A6. Since each lane yields 140 million single-end reads, a transcription profile (relative expression) of a sample can most times be made from sequences generated in one-third flowcell lane. For detection of splice variants or low-expressing genes, then up to one full lane of data may be required.

References :

1. Illumina, Inc. (2011) “RNAExpression_DataSheet.pdf” http://www.illumina.com/Documents/products/whitepapers/whitepaper_RNASeq_to_arrays_comparison.pdf
2. Shendure, Jay (2008). "The beginning of the end for microarrays?". *Nature Methods* 5 (7): 585–587.